

Curating Virtual Data Collections

A Virtual Collection synthesizes remote data and information resources related to a specific theme into a machine-actionable file of metadata that can be consumed by a variety of data applications to acquire those resources.

IN33B-1800
AGU Fall Meeting, 2015, San Francisco

Chris Lynnes, NASA/GSFC, Code 586, EOSDIS System Architect
Amanda Leon, National Snow and Ice Data Center
Hampapuram Ramapriyan, Science Systems and Applications, Inc
Vardis Tsonos, NASA/JPL
Chung-Lin Shie, NASA/GSFC
Zhong Liu, NASA/GSFC

Motivation:

Aid discovery of and access to data and information that pertain to a common theme but are dispersed and managed separately

Potential Beneficiaries:

End-Users: science teams, researchers, applications users, educators
Curators: science data providers, observing systems, journals

Brought to you by...the Virtual Collection Working Group

A NASA-chartered group to develop the Virtual Collection concept and specifications (expected Spring 2016)

Data can be grouped around a variety of themes, from research to applications to education, at varying levels of granularity

Themes	Examples	Data Resource Specification
Event Type	Volcanic Eruptions	datasets
Event Group	Mt. Etna Eruptions	datasets + spatial area
Event	Mt. Etna Eruption 2008-05-13 to 2009-07-06	specific data files / subsets
Research Area	Aerosols	datasets
Application	Agriculture	datasets + spatial area
Class Lab Exercise	"Shake and Bake" Week 4	specific data files / subsets
Published Paper	DOI: 10.1029/2010JB007906	specific data files / subsets
Field Experiment	OLYMPEX 2015-2016	datasets
Portal	Project portal	datasets
Official Report findings provenance	National Climate Assessment	datasets + spatial area + time range

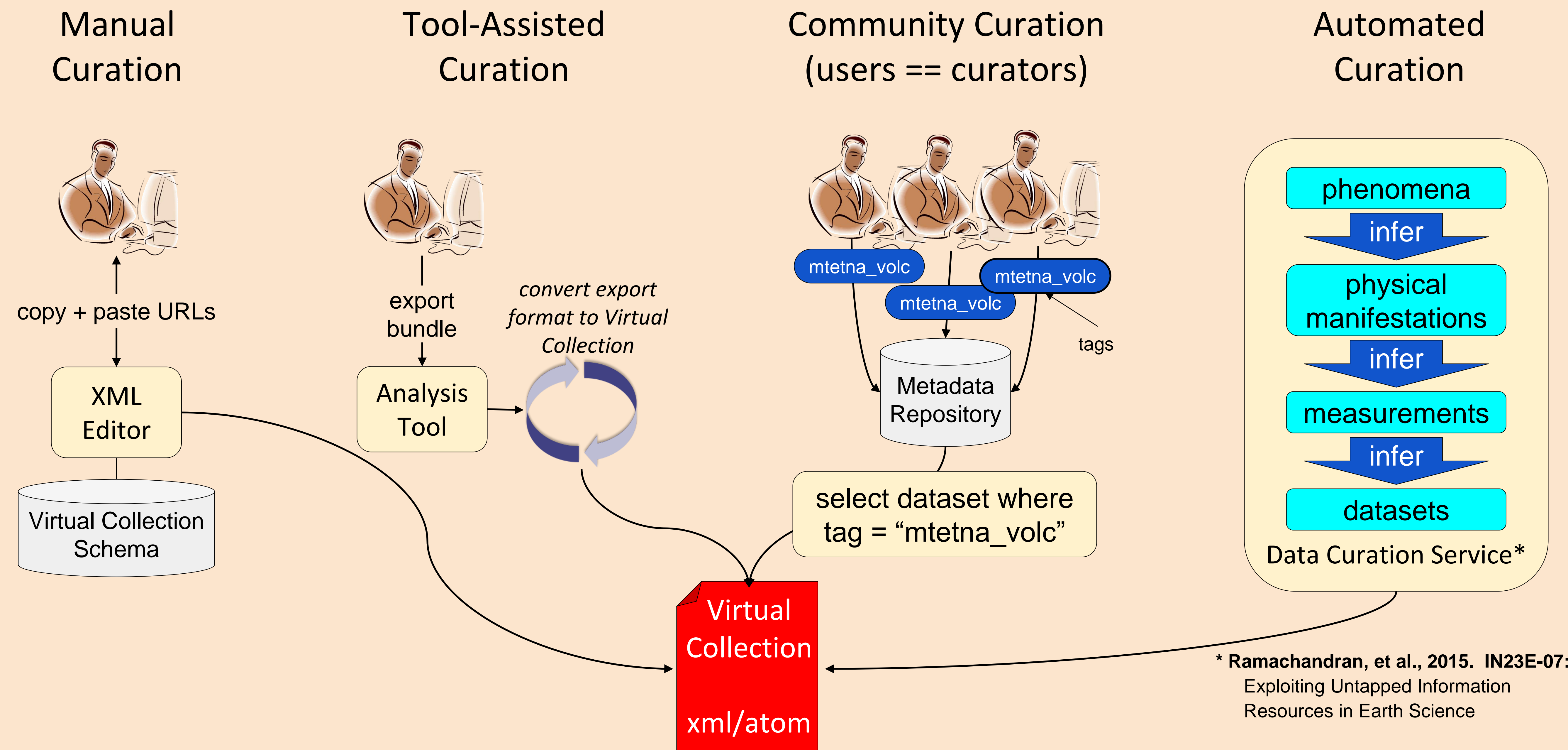
Virtual Collections also include URLs for related documents and information resources

Similarly, resources may have varying levels of granularity

Data Resource Type	Notional Examples
<i>Fully Qualified (Ready to use)</i>	<i>Ready-to-use URLs</i>
Dataset Directory	ftp://ladsweb.../.../MOD08_D3/
Static Data File	ftp://ladsweb.../MOD....hdf
Web Map Service GetMap request	http://.../...?...GetMap...
Web Coverage Service GetCoverage request	http://.../...?...GetCoverage...
OPeNDAP* netcdf response	http://.../...hdf.nc
<small>*Open-source Project for a Network Data Access Protocol</small>	
OPeNDAP service endpoint	http://.../...hdf
webification URL	http://.../.../*_L2StdND_*.h5/
Published Article	http://dx.doi.org/...
<i>Partially Qualified</i>	<i>URLs requiring more work to be usable...</i>
OpenSearch Query	http://.../...?...&bbox=-14,37,16,39&start=2008-05-09T00:00:00Z&end=2008-05-10T23:00:00Z&...
OpenSearch Description Document	http://.../.../..osdd.xml
Web Coverage Service DescribeCoverage request	http://.../...?...Describe Coverage...
Web Map Service DescribeLayer request	http://.../...?...Describe Layer...
Giovanni Workflow	http://.../...starttime=2008-05-01T00:00:00Z&endtime=2009-09-30T23:59:59Z&bbox=14,37,16,39...

Virtual means not having to copy everything (or anything) to one place

Curation Methods for Virtual Collections



Conversion to Application-Usable Form

